

# Complex-Value Recurrent Neural Networks for Global Optimization of Beamforming in Multi-Symbol MIMO Communication Systems

Danchi Jiang

School of Engineering, University of Tasmania, Private Bag 65, Hobart, TAS 7001, Australia.  
Email: danchi.jiang@utas.edu.au

**Abstract**—Multiple antennas at transmitter and receiver can be used to improve communication efficiency by canceling channel noises using the correlated information among the signals transmitted from different antennas. In this paper, a novel approach is proposed for this problem for another interesting case where multiple symbols are used to make the best use of the multiple antenna channel. Such an issue cannot be converted into a convex optimization problem. It can be considered as a generalization of the vector optimization problem on Grassmannian manifold to that one on a complex Stiefel manifold, which has not been well considered yet. The proposed algorithm is based on the gradient search on a complex Stiefel manifold of a non-convex problem to maximize the system signal to noise ratio. With appropriately defined Riemannian metric on this manifold, a neat formula has been developed for the gradient function. It is proved that the proposed algorithm converges to the global optimum. This algorithm can also be implemented into recurrent neural network to facilitate real-time computation. Its parallel structure can be realized using analog circuits. Furthermore, a modified gradient flow defined on the non-compact Stiefel manifold is also developed, which is robust against any initial condition error. The corresponding recurrent neural network is also discussed. Simulation experiments are included to demonstrate the advantages of the proposed algorithms.

## I. INTRODUCTION

Beamforming issues arise in many communication and signal processing areas such as radar, sonar, seismology, microphone array speech processing, etc. Please see [1] for some details and references therein. It is a critical problem in multi-input multi-output (MIMO) wireless communication channels. See [1]–[8] for a partial list of articles discussing such issues.

In particular, in MIMO channels multiple antennas are used to transmit messages to the receiving antenna array. The space difference among antennas can be utilized to achieving higher communication capacity than that in a single antenna channel. It is known that, the signal-to-noise ratio (SNR) is maximized in the principle eigenvector direction, which corresponding to the largest eigenvalue, of the channel gain matrix between transmitting antennas and receiving antennas, in the case where binary message is modulated using baseband signal transmitted through multiple antennas. As such, singular value decomposition method can be used to solve the beamforming problem in this simplest case where channel gain is completely known to both the transmitter and receiver, the channel is flat-fading, and signal to be transmitted is binary. For a brief discussion, please see [2].

More sophisticated techniques are required to take into account more practical considerations. In one case, channel information is unknown or not precisely known. The system needs to be adjusted adaptively according to channel information updated interactively between the transmitter and the receiver. For example, in [4] a simple algorithm is developed to tune the beamforming vector based on binary feedback from the receiver. In [1], [3], [5], [6], several methods are proposed to quantize and to adaptively adjust the beamforming vector based on finite rate feedback from receiver. In those works, the communication channel is not required to be reciprocal. That is, channel state information may not be observable at the transmitter and needs to be passed from the receiver in finite rate.

In another case, different constraints are included to reflect various restrictions in practical settings. In the award winning paper [7], a unified framework is proposed to minimize a cost function subject to transmitting power constraints for multicarrier communication. With UV decomposition of the beamforming vector, the original problem is converted into a convex optimization problem. Hence it can be readily solved.

Grassmannian manifold is used in [8] to represent the decision beamforming vector subject to equality norm constraint. The constrained optimization for beamforming is converted into an unconstrained one on this manifold. The advantages of this approach are two-fold: first, an elegant gradient formula can be obtained. Second, a clear insight of the optimal solution can be obtained, in comparison the matrix algebraic method such as SVD algorithm can only provide a procedural algorithm to obtain the numerical solution. Based on that, line packing problem is introduced for channel state information quantization.

In this paper, multiple symbols for MIMO communication are proposed to be used. By this way, not only the first principal eigenvector direction but also other eigenvector directions are utilized for beamforming to increase the capacity of this single channel. It is naturally to require in the system design that the interference between symbols is zero and the power for each symbol is equal. This incurs a nonlinear matrix equality constraint on the beamforming vector. As the result, it becomes very difficult, if it is still not impossible, to convert the beamforming design issue into a convex optimization problem. In fact, such a group of constraints is formulated into the complex Stiefel manifold, which is the high dimensional extension of Grassmannian manifold. A Riemannian metric induced from the ambient Euclidean space is used to calculate the gradient of the cost function. It is shown that, along the trajectory of the gradient flow, the cost function converges to its optimal and the decision vector converges to a number of eigenvectors corresponding to the largest singular values.

Following the seminal work of Hopfield and Tank [9], recurrent neural networks have been successfully applied to solve many classes of optimization problems in real time; see [10], [11] and references

cited therein for a few examples. Unlike the standard neural network that involves some parameters to be tuned during the learning phase, the principal idea of using recurrent neural network for optimization is to appropriately construct a neural network which represents the gradient direction of a given cost function and the states of the neural network represent decision variables, so that the network states converge to the optimal solution. The advantage of such networks are two-fold: first, it has a parallel structure and hence renders distributed computing; Second, it can be computed in real-time if implemented using analogue circuit or fast multi-processor computers.

Based on the analytic formula of gradient obtained on Stiefel manifold, a recurrent neural network can be designed. Its architecture is discussed. Furthermore, an alternative problem defined on a non-compact Stiefel manifold is proposed to accommodate the case where the initial Stiefel manifold. The resultant recurrent neural network is more robust.

This paper is organized into several sections. In the next section, the problem is introduced and carefully defined. Based on that, the gradient is calculated in the third section, where the convergence of the positive gradient flow is also established. The fourth section is dedicated to the discussion of recurrent neural network architecture. In the fifth section, an alternative problem formulation is proposed, which is not dependent on the Stiefel manifold. Simulation experiment results are included in the sixth section. A brief summary and discussion on the proposed method is organized in the last section to conclude this paper. Technical details for theoretical results are arranged in the appendix.

## II. PROBLEM FORMULATION

Consider a wireless system with multiple antennas at the transmitter and receiver. Assume that the communication channel information is completely known to simplify the analysis. The objective of beamforming is to choose appropriate transmitting weight, while the objective of antenna combining is to choose appropriate weight at receiver so that the received signal SNR is maximized.

In this paper we only discuss the beamforming at the transmitter. Such a channel can be illustrated

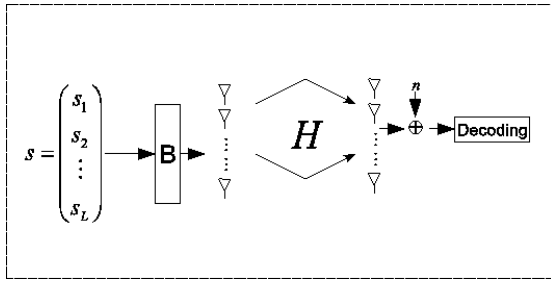


Fig. 1. MIMO Wireless System

using the diagram in Figure 1. In order to take the advantage of multiple antennas of the communication channel, we consider that the transmitter uses multisymbol instead of binary in source code modulation. In this case, let the transmitting weights (the beamforming matrix) be denoted  $B$ . The channel transfer matrix is denoted as  $H$ . The transmitted signal is  $s$  and the received signal is  $\bar{s}$ . The communication channel can be described as

$$\bar{s}(t) = HBs + n, \quad (1)$$

where  $n$  is the channel noise vectors. Without loss of generality, it is assumed to be Gaussian with zero mean and unit variance.  $s$  is a vector that represents signal sources with symbol length  $l$  sent to  $m$  antennas.  $B$  is a  $m \times l$  beamforming matrix. Assume the source signal is of unit power. It is expected that the resultant signals is of equal power and zero ISI before transmitting. Therefore,  $B^*B = I_l$ , where the upper index  $\star$  denotes the complex conjugate transpose of the original matrix. The following problem is formulated to maximize the signal power so that the SNR at the receiver can be maximized.

$$\begin{aligned} &\text{maximize: } J = \|HB\|^2, \\ &\text{subject to: } B^*B = I_l, \end{aligned} \quad (2)$$

where the matrix norm is Frebenuous norm. Let  $R = H^*H$ , the maximal SNR problem given in (2) can be further simplified as the following generalized Rayleigh quotient maximization problem:

$$\begin{aligned} &\text{maximize: } J = \text{tr}(X^*RX), \\ &\text{subject to: } X \in \mathbb{C}^{m \times l} \text{ and } X^*X = I_l. \end{aligned} \quad (3)$$

### III. CALCULATION OF GRADIENT AND PROPERTIES OF GRADIENT FLOWS

Defined the complex Stiefel manifold as follows:

$$St(l, m, \mathbb{C}) = \{X \in \mathbb{C}^{m \times l}, X^*X = I_l\}. \quad (4)$$

It is known that the complex Stiefel manifold is a compact manifold. See [12], [13] for some details. Its tangent space at  $X$  can be calculated as

$$T_X St(l, m, \mathbb{C}) = \{V \in \mathbb{C}^{m \times l} \mid V^*X + X^*V = 0_l\}. \quad (5)$$

Consider  $\mathbb{C}^{m \times l}$  as a  $2ml$  dimensional real vector space. It is easy to check the tangent space  $T_X St(l, m, \mathbb{C})$  is  $2ml - l(l + 1)$  dimensional one, noticing that the constraints consist of a group of  $l(l + 1)$  independent linear equations on it real and imaginary parts.

Define the following Riemannian metric on  $T_X St(l, m, \mathbb{C})$  as:

$$\begin{aligned} \langle\langle V_1, V_2 \rangle\rangle &= \text{tr}(V_1^*V_2 + V_2^*V_1) \\ &= \text{tr}(V_2V_1^* + V_1V_2^*). \end{aligned} \quad (6)$$

Such metric is induced from  $\mathbb{C}^{m \times l}$ . With this Riemannian metric, the gradient of the cost function in (3) can be calculated.

*Lemma 1:* The normal vector space of the tangent space at  $X$  of the Stiefel manifold is given by

$$T_X^\perp St(l, m, \mathbb{C}) = \{X\Lambda \mid \Lambda = \Lambda^* \in \mathbb{C}^{l \times l}\}. \quad (7)$$

Let the gradient of the cost function  $J$  on the complex Stiefel manifold be denoted as  $G(X)$ . Then,

- $\langle\langle G(X), V \rangle\rangle = dJ|_X(V), \forall V \in T_X St(l, m, \mathbb{C})$ .
- $G(X) \in T_X St(l, m, \mathbb{C})$ .

The following results list the gradient formula and the convergence property of the gradient flow.

*Theorem 1:* Consider the cost function  $J$  given in (3) defined on the complex Stiefel manifold given by (4). With respect to the induced Riemannian metric in (6), the gradient can be calculated as

$$\text{grad}J(X) = (I_m - XX^*)RX. \quad (8)$$

Furthermore, the trajectory of the positive gradient flow defined by the following equation

$$\dot{X} = \text{grad}J(X), \quad (9)$$

converges to the global maximum of the cost function.

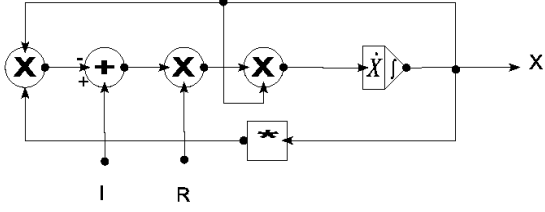


Fig. 2. The architecture of the proposed recurrent neural network.

#### IV. THE PROPOSED RECURRENT NEURAL NETWORKS

A recurrent neural network can be designed based on the gradient flow given by (9). Its architecture can be illustrated in the following diagram:

In Figure 2, the state of the recurrent neural network is a complex matrix. The operators are matrix operators to simplify the diagram. The star symbol “\*” represents the complex conjugate transpose operator. In practical implementation, complex numbers can either be separated into real and imaginary part, or simply be the state of the electronic circuits. In this paper, detailed circuit design for neural networks is not discussed. Interested reader may refer to [14] for more details.

#### V. ROBUST RECURRENT NEURAL NETWORK

In this section, sensitivity of the optimal solution is considered for the cases where there are initial condition error or the channel transfer matrix error. Because any intermediate state of a neural network can be considered as an initial state for time afterwards, this consideration can also take into account the deviation of state from the precise trajectory due to small hardware or computing precision errors.

For the proposed recurrent neural network representing the gradient flow of the cost function, the state matrix remains on the Stiefel manifold if it starts from an initial state on that manifold. However, if the initial state is not on the Stiefel manifold, more specifically, its column vectors are not orthonormal to each other, the state of the proposed recurrent neural network is not guaranteed to converge to this manifold. A simulation experiment is included in the next section to demonstrate this sensitivity. As such, the recurrent neural network

proposed in Section IV is not ready to use yet. To remedy this, the following modified problem is proposed.

$$\begin{aligned} &\text{maximize: } J = \text{tr}\{X^*RX(X^*X)^{-1}\}, \\ &\text{subject to: } X \in \mathbb{C}^{m \times l} \text{ and } \text{rank}(X^*X) = l. \end{aligned} \quad (10)$$

Since the maximal solution to the problem given in (3) is an admissible solution to the problem defined in (10), the maximum of the cost function in (10) is not less than that of (3). On the other hand, if the maximal solution to (10) is  $\hat{X}$ , the matrix  $X(X^*X)^{-1/2}$  is also an admissible solution to the problem defined in (3). As such, the original optimization problem given by (3) and the modified problem in (10) is equivalent. However, the admissible set of (10) is an open set.

Define the non-compact complex Stiefel manifold as

$$\text{ST}(l, m, \mathbb{C}) = \{X \in \mathbb{C}^{m \times l} \mid \text{rank}X = l\}, \quad (11)$$

and define the Riemannian metric on this manifold as

$$\begin{aligned} \ll U, V \gg: &= \text{tr}(U(X^*X)^{-1}V^* + V(X^*X)^{-1}U^*), \\ &\forall U, V \in T_X\text{ST}(l, m, \mathbb{C}). \end{aligned} \quad (12)$$

It can be computed that

$$\begin{aligned} &DJ_X(V) \\ &= \text{tr}\{V^*RX(X^*X)^{-1} + X^*RV(X^*X)^{-1} \\ &\quad - X^*RX(X^*X)^{-1}[V^*X + X^*V](X^*X)^{-1}\} \\ &= \ll (I - X(X^*X)^{-1}X^*)RX, V \gg. \end{aligned} \quad (13)$$

As such, the positive gradient flow can be defined as

$$\dot{X} = (I - X(X^*X)^{-1}X^*)RX. \quad (14)$$

It can be observed that this gradient flow coincide with the one defined in (9) if  $X$  is on the complex Stiefel manifold. Furthermore, direct computation shows that

$$\frac{d}{dt}(X^*X) = 0, \quad (15)$$

along the gradient flow given in (14). Therefore, any initial matrix on the complex Stiefel manifold remains on it along the trajectory of the gradient flow. Hence this gradient flow is robust with respect to the initial condition error.

The architecture of the corresponding recurrent neural network can be illustrated using the diagram in Figure 3.

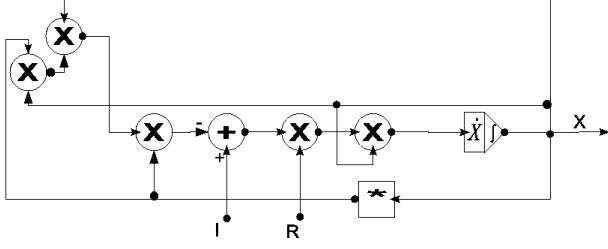


Fig. 3. The architecture of the robust recurrent neural network.

## VI. SIMULATION EXPERIMENTS

In this section, some experiments data are shown to illustrate the proposed recurrent neural networks. For the sake of cost, the experiments have been conducted using Matlab and Simulink. However, they can represent the state of the neural networks developed in this paper precise enough by adjusting the numerical precision of the simulation tool.

Consider a simple case where the beamforming matrix is  $3 \times 2$ . Even though this is very small in size, the simulation outcomes can already show these convergence properties of the recurrent neural network discussed in previous sections. The initial beamforming matrix is generated using a random number generator. The channel transfer matrix is also randomly generated. In our test, they are given by the following matrices.

$$X_0 = \begin{pmatrix} 0.9342 & 0.8729 \\ 0.2644 & 0.2379 \\ 0.1603 & 0.6458 \end{pmatrix},$$

$$R = \begin{pmatrix} 5.8836 & 3.9969 & 2.7023 \\ 3.9969 & 5.0093 & 1.5566 \\ 2.7023 & 1.5566 & 1.3495 \end{pmatrix}.$$

To this initial matrix, the constraint  $X^*X = I_2$  is not satisfied. The following Figures 4, and 5 show the lognorm of the gradient and the cost function along the trajectory of the proposed recurrent neural network given in (9). In particular, it can be seen that the cost function is not monotonically increasing, which is not what we expect for a gradient flow. This is due to the violation of the constraint  $X^*X = I_2$ .

In Figures 6 and 7, the same initial condition and the the channel transfer matrix are applied to the modified recurrent network depicted in Figure

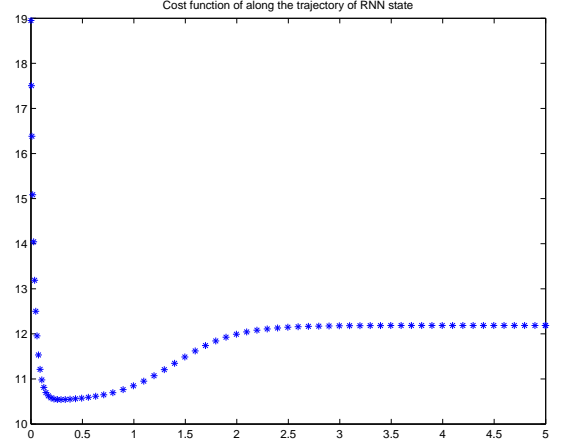


Fig. 4. The cost function along the trajectory of the proposed recurrent neural network.

3. The results demonstrate a good convergence property. Not only the gradient converges to zero, but also the cost function increases monotonically toward the global maximum.

However, if the orthonormal constraint  $X^*X = I_2$  on initial state is satisfied, the proposed recurrent neural network can still demonstrate a very good convergence property. To this end, let the initial state be  $X_0(X_0^*X_0)^{-1/2}$ . It can be easily checked that this initial state is orthonormal. Simulate the proposed recurrent neural network given in Figure 2 with this initial state and the channel transfer matrix  $R$  generated before. The resultant cost function increasing monotonically. See Figures 8 and 9 for the curve of cost function and the log norm of the gradient. Even though the norm of gradient is not monotonically decreasing, it converges to zero

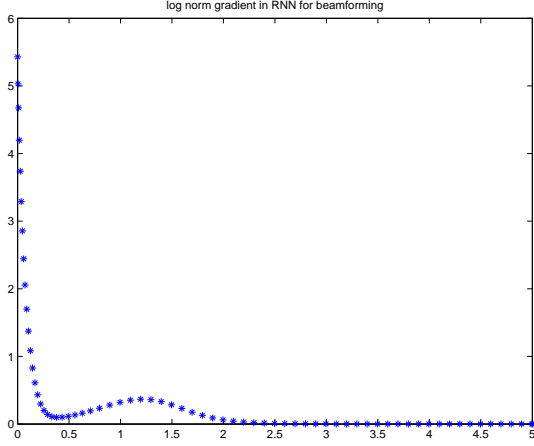


Fig. 5. The log norm function of the gradient along the trajectory of the proposed recurrent neural network.

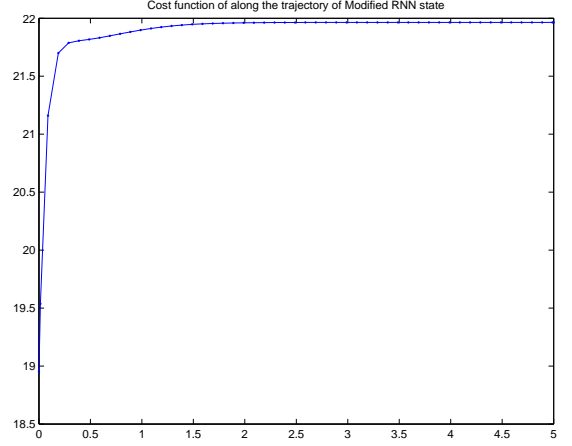


Fig. 6. The cost function along the trajectory of the modified recurrent neural network.

quickly.

## VII. CONCLUSION

In this paper, two novel recurrent neural networks are developed based on the positive gradient flows of the generalized Rayleigh quotient on a complex Stiefel manifold or on a non-compact complex Stiefel manifold. The advantage of such networks are two-fold: first, it has a parallel structure and hence renders distributed computing; Second, it can be computed in real-time if implemented using analogue circuit or fast multi-processor computers. It is shown the gradient flow on the complex Stiefel manifold will converge to the global optimal solution of the formulated multi-symbol MIMO beamforming problem. The generalized Rayleigh quotient on a non-compact Stiefel manifold is equivalent to that defined on the complex Stiefel manifold

with a simple state transformation. However, the orthonormal constraint in such case is removed. Both technical analysis and simulation experiments are included to demonstrate the advantages of the proposed approach.

## APPENDIX: TECHNICAL ANALYSIS

**Proof of Lemma 1.** This is because for any  $V \in T_X St(l, m, \mathbb{C})$ , there holds

$$\begin{aligned} \ll V, X\Lambda \gg &= \text{tr}(V^* X\Lambda + \Lambda^* X^* V) \\ &= \text{tr}\{\Lambda[V^* X + X^* V]\} = 0. \end{aligned}$$

Which show that

$$\{X\Lambda \mid \Lambda = \Lambda^* \in \mathbb{C}^{l \times l}\} \subseteq T_X^\perp St(l, m, \mathbb{C}).$$

On the other hand, the dimension of the set  $\{X\Lambda \mid \Lambda = \Lambda^* \in \mathbb{C}^{l \times l}\}$  can be calculated as  $2ml -$

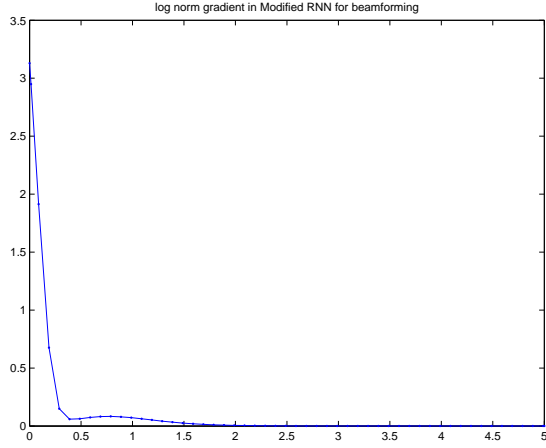


Fig. 7. The log norm function of the gradient along the trajectory of the modified recurrent neural network.

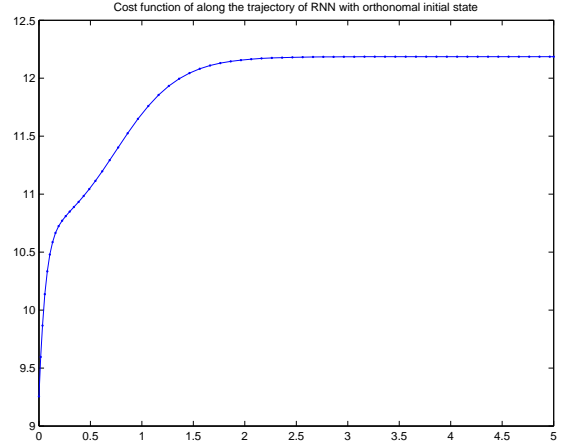


Fig. 8. The cost function along the trajectory of the proposed recurrent neural network with orthonormal initial state.

$\dim T_X St(l, m, \mathbb{C})$ , which is exactly the same as the dimension of  $T_X^\perp St(l, m, \mathbb{C})$ . As such, two complex vector sets (but considered as real vector spaces) are identical. ■

**Proof of Theorem 1.** From the first condition of the gradient we know that  $G(X) - RX \in T_X^\perp St(l, m, \mathbb{C})$ . As such, the gradient must be in the form of  $RX + X\Lambda$ . On the other hand, the second condition leads to

$$\Lambda = -X^*RX.$$

Therefore,

$$\text{grad}J(X) = (I_m - XX^*)RX.$$

Furthermore, consider the second order derivative

$$D^2J_X(V, V) = 2\text{tr}(V^*RV).$$

Since  $R$  is positive semi-definite,  $D^2J_X(V, V) = 0$  leads to  $RV = 0$ . As such, the Hessian of the cost function is zero can only be possible where the gradient of the cost is also zero. Hence, the cost function is a Morse-Bott function. From the convergence property of a Morse-Bott function, see [15], it is guaranteed that the gradient flow given by

$$\dot{X} = \text{grad}J_X = (I_m - XX^*)RX. \quad (16)$$

defines a trajectory convergent to the global optimal solution. ■

## REFERENCES

- [1] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2257–2269, 2003.
- [2] S. Haykin and M. Moher, *Modern Wireless Communications*. U. S. A.: Pearson Prentice Hall, 2005.

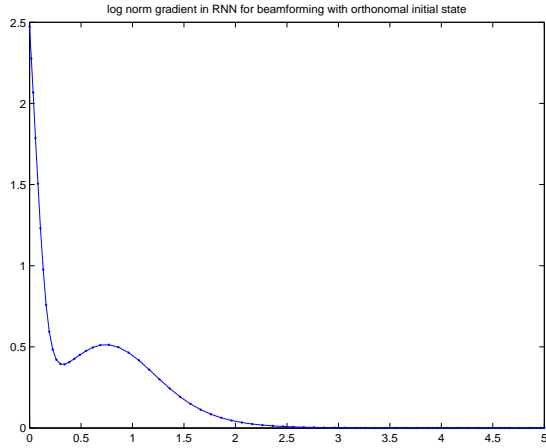


Fig. 9. The log norm function of the gradient along the trajectory of the proposed recurrent neural network with orthonormal initial state.

- [3] B. Mondal and R. W. J. Health, "Channel adaptive quantization for limited feedback mimo beamforming systems," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4717–4729, 2006.
- [4] B. C. Banister and R. Z. James, "A simple gradient sign algorithm for transmit antenna weight adaptation with feedback," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1156–1171, 2003.
- [5] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Transactions on information theory*, vol. 49, pp. 2562–2579, 2003.
- [6] R. B. D. Roh, June Chul, "Transmit beamforming in multiple-antenna systems with finite rate feedback: a vq-based approach," *IEEE Transactions on Information Theory*, vol. 52, pp. 1101–1112, 2006.
- [7] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: a unified framework for convex optimization," *IEEE Transactions on signal processing*, vol. 51, pp. 2381–2401, 2003.
- [8] D. J. LOve, R. W. J. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Transactions on Information Theory*, vol. 49, pp. 2735–2747, 2003.
- [9] D. W. T. . J. J. Hopfield, "Simple neural optimization networks: An a/d converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. Circuits and Systems*, vol. 33, no. 5, pp. 533–541, 1986.
- [10] D. Jiang and J. Wang, "A recurrent neural network for real-time semidefinite programming," *IEEE Transactions on Neural Networks*, vol. 10, pp. 81–93, 1999.
- [11] —, "Augmented gradient flows for on-line robust pole assignment via state and output feedback," *Automatica*, vol. 38, pp. 279–286, 2002.
- [12] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, pp. 635–650, 2002.
- [13] S. Fiori, "Non-linear complex-valued extensions of hebbian learning: An essay," *Neural Computation*, vol. 17, pp. 779–838, 2005.
- [14] C. A. Mead, *Analog VLSI and Neural Systems*. United States of America: Addison-Wesley, 1989.
- [15] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, 2nd ed. London: Springer-Verlag, 1996.